

## Ziusudra, de la nomenclature à l'informatique: l'exemple des Amphibiens<sup>1</sup>

David GÉRARD\*, Régine VIGNES-LEBBE\*\* & Alain DUBOIS\*

\* Reptiles et Amphibiens, USM 0602 Taxonomie & Collections,  
Département Systématique & Évolution, Muséum national d'Histoire naturelle,  
Case postale 30, 25 rue Cuvier, 75005 Paris, France

<dagerard@mnhn.fr>, <adubois@mnhn.fr>

\*\* Laboratoire Informatique et Systematique (LIS),  
UMR 5143 Paleobiodiversite & Paléoenvironnements, Equipe Systematique,  
Recherche Informatique et Structuration des Cladogrammes Université Pierre & Marie Curie,  
12 rue Cuvier, esc B, 2e étage, 75005 Paris, France  
<vignes@ccr.jussieu.fr>

**The *International Code on Zoological Nomenclature* deals with nomina (scientific names) of taxa of ranks between subspecies and superfamily, i.e. so-called "lower nomenclature". No Rules exist for "higher nomenclature", which is therefore informal, instable and confusing. A recent proposal of Rules for higher nomenclature relies on the original contents and limits of taxa for which the nomina were proposed, as compared with the current contents and limits of taxa in a given taxonomy. Furthermore, in order not to upset two and a half centuries of tradition, nomina in large use, especially outside the specialised field of systematics, need to be conserved. This work presents a formalization for the representation of nomenclatural data and the automation of the application of the proposed nomenclatural Rules to these data. A prototype of software was developed. Its architecture is based on: (1) a database which includes the nomenclatural information needed for the formalization proposed; (2) a program at file server end, which queries the database and implements the proposed rules; (3) a web interface which allows to operate the program. The prototype was developed first on the example of the nomina of the class Amphibia**

### INTRODUCTION

Après 250 ans de taxinomie et de nomenclature, où en sommes-nous de la gestion des noms scientifiques? Doit-on écrire INSECTA ou HEXAPODA, URODELA ou CALDATA? Doit-on se préoccuper de la distinction "nomenclature supérieure - nomenclature inférieure"? Le travail présenté dans cet article constitue un élément de réponse à ces deux questions.

<sup>1</sup> Communication présentée lors de la table ronde "Systematique et diversité des Amphibiens" organisée par l'ISSCA et tenue à l'Université d'Angers (France) le 3 février 2006

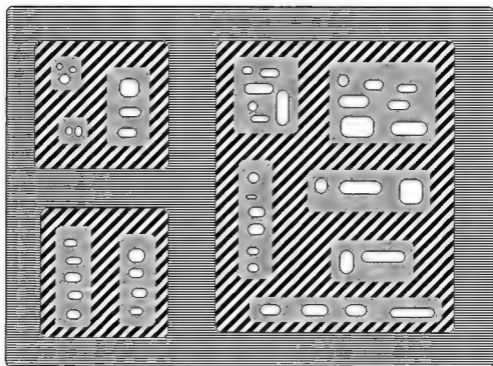


Fig. 1 Hiérarchie mathématique. Chaque sous-ensemble est contenu dans un plus grand ensemble, lui-même contenu dans un plus grand ensemble. On parle de hiérarchie mathématique car la notion d'inclusion se superpose à la notion d'ordre.

La taxinomie (ou plus exactement la *macrotaxinomie*, MAYR & ASHLOCK, 1991) vise à établir une classification systématique des êtres vivants (DARLU & TASSY, 1993). La nomenclature, quant à elle, a pour fonction de donner des noms scientifiques ou *nomina* (DUBOIS, 2000) aux taxons, de manière telle que tous les utilisateurs de ces *nomina* puissent parler sans ambiguïté de la même chose (ANONYME, 1999). Taxinomie et nomenclature ont une structure hiérarchique (fig. 1), comportant un ensemble de rangs successifs, dont sept principaux (regne, embranchement, classe, ordre, famille, genre et espèce) et divers rangs additionnels, pour la plupart désignés par des préfixes (sous-espèce, super-famille, etc.). Ces rangs permettent de formaliser une structuration des relations de subordination (ou d'inclusion) entre taxons, traduites dans une hiérarchie entre *nomina* (KNOX, 1998). Les deux disciplines taxinomique et nomenclature, bien qu'intimement liées, sont distinctes. Le but de la nomenclature est d'accompagner la taxinomie, non de la figer (ANONYME, 1999).

Au sein de la nomenclature zoologique, on peut définir deux sous-ensembles : (1) la nomenclature dite supérieure, qui porte sur les noms scientifiques appartenant à l'intervalle de rangs allant du regne à l'ordre (*nomina* de la "série-classe", sensu DUBOIS, 2000, 2005a-c, 2006a-c); et (2) la nomenclature dite inférieure, qui porte sur les noms scientifiques apparte-

nant à l'intervalle de rangs allant de la famille à l'espèce (nomina de la "série-famille", de la "série-genre" et de la "série-espèce", sensu DUBOIS, 2000, 2005a-c, 2006a-c)

La nomenclature zoologique inférieure repose sur le *Code International de Nomenclature Zoologique* (ANONYME, 1999; dénommé "le Code" ci-dessous), qui permet de régir les noms scientifiques selon des Règles précises et assez souvent contraignantes mais nécessaires. Pour faire face aux problèmes qui se posent parfois, statuer en cas de litiges, et permettre à cette nomenclature d'évoluer, il existe une instance, la Commission Internationale pour la Nomenclature Zoologique (dénommée "la Commission" ci-dessous). Celle-ci publie un périodique trimestriel, le *Bulletin of Zoological Nomenclature*, destiné à discuter des problèmes et des propositions soumises par les zoologistes du monde entier, et à rendre publiques les décisions de la Commission.

La nomenclature supérieure n'est actuellement pas régie par des Règles. Certains critères consensuels sont parfois utilisés pour choisir entre plusieurs nomina potentiels pour un taxon, mais sans Règle précise formalisée, si bien que le résultat de leur application ne peut être ni objectif ni nécessairement répétable. Ces critères de choix d'un nomen reposent sur: (1) la priorité d'ancienneté entre deux nomina, sans tenir le plus souvent compte de l'évolution du contenu (extension) du taxon correspondant, (2) la nature plus ou moins appropriée du nomen, par exemple, pour parler d'un animal à six pattes, le nomen **HEXAPODA** semble plutôt adapté, (3) l'emploi du grec et/ou du latin pour former le nomen, (4) l'"usage", mais sans définition précise de ce qui est un usage significatif.

Ces critères vagues, employés en nomenclature supérieure, ne sont pas cohérents avec le *Code*, qui s'applique aux nomina de taxons inférieurs. Il existe donc actuellement deux fonctionnements différents au sein de l'ensemble de la nomenclature zoologique, ce qui n'est pas satisfaisant pour l'esprit et porte en germe une remise en cause fondamentale du *Code*, celui-ci n'ayant pas une fonction universelle pour tous les nomina zoologiques, contrairement à d'autres systèmes alternatifs récemment proposés, comme le *Phylocode* (LAURIN & CANTINO, 2004). D'autre part, la généralisation des analyses cladistiques induit de nos jours une multiplication des nomina de taxons supérieurs, ou encore le mauvais emploi des nomina existants. De ce double constat se dégage la nécessité de formaliser la nomenclature supérieure.

## UN NOUVEAU SYSTÈME POUR LA NOMENCLATURE SUPÉRIEURE

Les Règles récemment proposées pour la nomenclature supérieure (DUBOIS, 2004, 2005a-d, 2006a,c) se basent sur les mêmes principes fondamentaux que la nomenclature inférieure, pour assurer logique et cohérence à l'ensemble de la nomenclature zoologique. Comme dans le cas des nomina régis par le *Code*, on parlera en nomenclature supérieure de: (1) *disponibilité* du nomen (terme "latinisé" publié sur un support durable et en vertu de Règles précises comme le demande le *Code* pour les nomina inférieurs); (2) *attribution* d'un nomen à un taxon, qui s'effectue par le biais d'un "type porte-nom" ou *onomatophore* (SIMPSON, 1940, 1961, DUBOIS & ÖHLER, 1997, DUBOIS, 2005c); (3) *validité* du nomen pour désigner un taxon précis en fonction de la priorité entre d'éventuels synonymes (objectifs ou subjectifs) ou homonymes.

En raison des particularités de la nomenclature supérieure, l'onomatophore ne correspondra plus à un spécimen ou un nomen mais à une liste de nomina de genres, inclus dans l'extension du taxon pour lequel il a été créé. Un autre concept est employé, celui de "protection" d'un nomen, proche de la notion d'usage pré-citée mais plus objectif. Cela revient à privilégier un synonyme par rapport à un autre, même s'il n'est pas prioritaire sur le plan de son antériorité mais parce qu'il est *significativement employé en dehors du domaine de la systématique*. Contrairement à la pratique récente de la Commission, qui a alimenté ces dernières années des débats animés, la protection ici relève de critères précis (DUBOIS, 2005b-d). En effet, la formule en italiques ci-dessus est définie précisément comme suit. (1) soit dans les *titres* d'au moins vingt-cinq livres non purement systématiques, écrits par vingt-cinq auteurs-indépendants (c'est-à-dire n'ayant jamais publié ensemble sur le taxon concerné) et publiés dans au moins dix pays différents après le 31 décembre 1899; (2) soit dans les *titres* d'au moins cent publications (livres, chapitres de livres ou articles de périodiques) non purement systématiques, écrites par cent auteurs-indépendants et publiés dans au moins dix pays différents après le 31 décembre 1899.

Une fois ces bases de nomenclature mises en place, la proposition originale de DUBOIS (2004, 2005a-d, 2006a,c) repose sur la comparaison entre la taxinomie originale ou *protaxinomie* (c'est-à-dire le contenu et les limites du taxon auquel se rapporte un nomen lors de sa première publication) et une taxinomie choisie. En fonction de sa pensée taxinomique, chaque chercheur peut désigner comme fonctionnelle une taxinomie différente, c'est-à-dire une *ergotaxinomie* donnée (DU BOIS, 2005c). La comparaison entre protaxinomie et ergotaxinomie peut donner lieu à trois cas de figures.

Considérons tout d'abord une ergotaxinomie actuelle représentée par la figure 2, et imaginons qu'un nomen N1 fut initialement proposé pour un taxon incluant les genres G1 et G4. Par rapport à l'ergotaxinomie considérée, le nomen N1 s'applique potentiellement à deux taxons de rangs supérieurs, T1 et T2, qui tous deux incluent les genres G1 et G4. Pour trancher entre les deux, en l'absence d'autres informations, le nœud le moins inclusif sera retenu, et le nomen N1 devra donc être attribué au taxon T2.

Toujours par rapport à la même ergotaxinomie de la figure 2, considérons maintenant un nomen N2, créé pour un taxon incluant les genres G2 et G6, et excluant explicitement le genre G7. Dans l'ergotaxinomie considérée, un seul taxon, T2, correspond à cette double caractérisation : c'est donc sans ambiguïté à ce taxon et lui seul que s'appliquera le nomen N2.

Considérons enfin un nomen N3 créé pour un taxon incluant les genres G2 et G8, et excluant explicitement le genre G5. Aucun taxon dans l'ergotaxinomie de la figure 2 ne correspond à cette caractérisation, puisque le taxon T1, qui inclut les genres G2 et G8, inclut également le genre G5. Dans un tel cas, on ne prendra pas en compte l'information apportée par l'exclusion d'un genre (qui est devenue incompatible avec l'ergotaxinomie actuelle), et on ne considérera que l'information apportée par la liste des genres inclus, ce qui nous ramène dans la même situation que pour le nomen N1 : ici le nomen N3 s'appliquera au taxon T1.

Les Règles simples présentées graphiquement ci-dessus et de manière plus détaillée ailleurs (DU BOIS, 2005b,d, 2006a,c) permettent l'*attribution* automatique et non-ambiguë de tout nomen de nomenclature supérieure à un taxon supérieur actuellement reconnu dans le cadre d'une ergotaxinomie donnée. Mais ceci n'implique pas que celui-ci soit le nomen *valide* du taxon à l'heure actuelle. En effet, en vertu de ces Règles, plusieurs nomina distincts peuvent

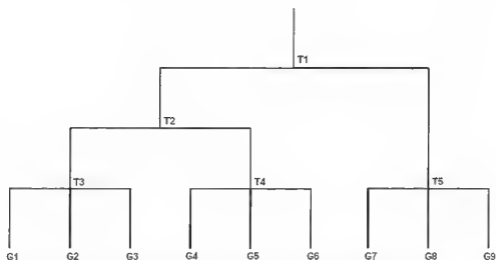


Fig 2 Arbre de taxons. Dans cet exemple, on considérera arbitrairement que les taxons de rang inférieur (G1 à G9) sont des genres, regroupés dans des taxons supérieurs (T1 à T5) de plusieurs rangs distincts. Le nomen N1 fut initialement proposé pour un taxon incluant les genres G1 et G4. Le nomen N2 fut initialement proposé pour un taxon incluant les genres G2 et G6 et excluant le genre G7. Le nomen N3 fut initialement proposé pour un taxon incluant les genres G2 et G8 et excluant le genre G5.

s'appliquer au même taxon (c'est le cas dans l'exemple ci-dessus des nomina N1 et N2). Dans un tel cas, comment trancher? Les Règles proposées font la distinction entre deux situations: (1) soit un nomen est *d'usage significatif*, aussi il sera protégé et validé, (2) soit ce nomen n'est pas *d'un usage significatif*, et il n'y a pas lieu de le protéger, dans un tel cas, comme en nomenclature inférieure, c'est la simple *priority de publication* qui permettra de déterminer le nomen à retenir comme valide pour le taxon.

A ces Règles simples doivent être adjointes diverses Règles annexes, destinées à gérer un certain nombre de cas particuliers et de problèmes spécifiques qui se posent en nomenclature supérieure en raison du fait que cette nomenclature est restée pendant 250 ans hors de tout contexte formalisé, et que les nomina des taxons supérieurs s'appliquent chacun à un seul taxon et non pas à une série de taxons coordonnés, comme dans la nomenclature inférieure où existe la Règle de Coordination (ANONYME, 1999). Les Règles proposées sont ainsi nécessairement complexes dans le détail (DUBOIS, 2006a), mais elles ont l'avantage de permettre de concilier la tradition centenaire d'emploi de certains nomina (comme *AMPHIBIA* ou *MOLLUSCA*) dans des milliers de publications non spécialisées en systématique, avec un besoin de rigueur, d'universalité et d'automatisme dans l'emploi de tous les autres nomina inconnus des non-spécialistes, ou très récemment créés. A cet égard, ces Règles sont certainement nettement préférables à un autre système récemment proposé pour la nomenclature supérieure (ALONSO-ZARAZAGA, 2005), qui envisage de remplacer tous les nomina, y compris les plus connus, par un système de nomina nouveaux suivant une Règle de Coordination, cette proposition catastrophique, si elle devait être suivie par la Commission, contribuerait sans nul

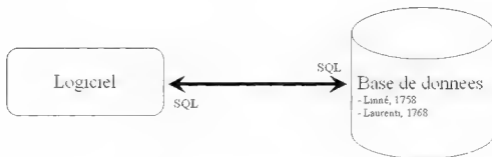


Fig. 3. – Architecture actuelle du prototype.

doute à éloigner du *Code* de nombreux utilisateurs des nomina zoologiques, pour les jeter dans les bras du *Phylocode* (DUBOIS, 2005c) ou d'un autre système alternatif.

## UNE PROPOSITION D'INFORMATISATION

### PRINCIPE

Un prototype nommé *Ziusudra*<sup>2</sup> (fig. 3) a été développé pour automatiser la recherche du nomen valide d'un taxon supérieur selon les Règles ci-dessus. Son objectif est donc de permettre à tout utilisateur de déterminer automatiquement quel est le nomen de la série-classe à utiliser aujourd'hui parmi les nomina disponibles, en fonction du contenu du taxon pour lequel chaque nomen a été créé. Décomposons ces différents aspects: (1) la mise à disposition d'un service sur internet permettra à tout utilisateur d'avoir accès à la même information, automatiquement et rapidement; (2) selon la formalisation de la nomenclature supérieure proposée, chaque nomen pourra être validé ou rejeté selon une succession d'étapes explicitement définies, (3) le service comportera un stockage d'informations nomenclaturales et taxinomiques, ce qui s'est traduit dans le cadre de ce travail par la mise en place d'une base de données. À l'heure actuelle, le prototype informatique décrit ci-dessous n'est opérationnel qu'avec une seule ergotaxinomie, mais pour pouvoir respecter la liberté de la pensée taxinomique un nombre indéfini d'ergotaxinomies pourront ultérieurement être comparées avec la taxinomie originale.

Développé en PERL (CHRISTIANSEN & TORRINGTON, 1999; WALL et al., 2001), déployé sur un serveur à l'aide du système de gestion de base de données PostgreSQL (WORSLEY & DRAKI,

2. Pourquoi *Ziusudra* (DUBOIS, 2006a)? La lecture de mythes anciens nous révèle que Noé (Noah tel qu'il apparaît dans la Bible) n'est pas le seul nom qui existe pour nommer celui qui a rassemblé une paire de chaque espèce vivante au sein de son arche. Sont à ce titre disponibles le nom hébreux Noakh, le nom grec Deucahon, le nom persan Yima, les noms indus Manu et Satyawrata, et le nom babylonien Utnaphistim. Mais le plus ancien, et qui aurait par conséquent la priorité, est le nom sumérien *Ziusudra*. Or s'il a pu réunir sur son arche un couple de chaque espèce, *Ziusudra* a dû développer une taxinomie pour les identifier, et probablement une nomenclature pour les nommer.

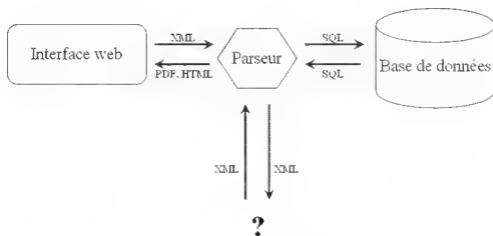


Fig. 4. – Architecture prévue du logiciel

2002), le service proposé permet à toute personne disposant d'un poste informatique connecté à internet d'interroger la base de données et de voir le résultat obtenu dans le cas de l'application des Règles proposées. Ci-dessous nous considérerons, à titre d'exemple, les deux premières publications concernant la taxinomie du groupe des Amphibiens (LINNAEUS, 1758; LAURENTI, 1768).

Cette architecture doit évoluer vers le schéma présenté en figure 4 pour s'enrichir de nouvelles possibilités. L'interface web communiquera donc en XML avec la base de données via un traducteur. Elle enverra au serveur une requête au format XML. Le traducteur interprétera cette requête, interrogera la base de données et renverra à l'interface web, donc au navigateur, une fiche-résultat en format XML.

Le parseur forme alors le coeur du logiciel car il permettra d'enrichir la base de données, ou simplement d'interroger la base, autrement que par l'interface mise en place. Pour que les résultats soient pertinents, il est primordial de faciliter la mise à jour de la base de données. La taxinomie est une science vivante faite d'hypothèses, plus ou moins stable selon les groupes mais qui reste en évolution continue. Le maintien à jour des informations nécessite alors une main d'œuvre qui fait le plus souvent défaut. Face à ce problème, ce logiciel se veut à long terme autonome dans le sens où il corrigera les informations présentes dans la base de données en fonction des nouvelles publications saisies. Dans un premier temps, un contrôle assuré par un ou des spécialistes sera nécessaire pour vérifier que l'automatisation des Règles produit un résultat cohérent, nomenclaturalement et taxinomiquement. La seule mise à jour nécessaire sera ensuite l'alimentation de la base de données en nouvelles publications. Celle-ci sera décentralisée et proposera une interface en ligne. Cette interface d'édition, déployée sur un serveur et accessible de tout poste informatique connecté à internet, ne pourra en revanche pas être utilisable par tout utilisateur, mais plutôt par des spécialistes qui s'authentifieront, ce qui permettra un suivi des saisies.

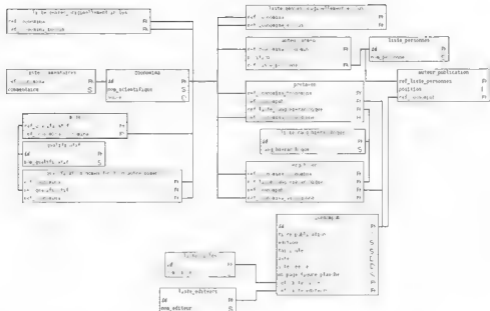


Fig 5 Schéma de la base de données. D, champ de type date, FK, foreign key ou référent, qui permet de lier une information d'une table a une information d'une autre table, I, champ integer ou entier; PK, primary key ou clé primaire, unique pour chaque entrée au sein de la table considérée; S, champ de type string ou chaîne de caractères.

#### SCHÉMA DE LA BASE DE DONNÉES

La base de données repose sur un schéma (fig 5), inspiré des travaux de BREINDSOHN (1995, 1997; BREINDSOHN et al., 1999) et de ZHONG et al. (1996, 1999). Les quatre tables centrales de ce schéma sont. (1) "zoonomina", où sont stockés les nomina et leurs dates de création; (2) "zoonompub", où sont stockées les publications, décrites par un certain nombre d'attributs ou colonnes, (3) "ergotaxon", où est stockée l'ergotaxinomie de référence (en l'occurrence, TRUB & CLOUTIER, 1991; TUDGE 2000, LECOINTRE & LI GU YADIR, 2001); (4) "protaxon", où est stockée la taxinomie originale.

La hiérarchie entre taxons est ici stockée en liant le taxon qui porte le nomen *n* au taxon surordonné qui porte le nomen *N*. Les tables "auteur\_nomen" et "auteur\_publication" sont ce qu'on appelle des tables de croisement qui permettent de restituer les relations de type "*n* à *m*" (GRAY et al., 2000). Elles permettent de lier plusieurs personnes (les noms de personnes étant stockés dans la table "personne") à une publication ou à un nomen et réciproquement, tout en conservant la position de l'auteur lorsqu'il y en a plusieurs. Les tables "liste\_genes\_originellement\_inclus" et "liste\_genes\_originellement\_exclus" permettent de stocker le contenu et les limites du protaxon pour lequel fut créé le nomen considéré. Les tables "rang\_hierarchique", "qualificatif", "personne", "liste\_editors" et "liste\_villes" permettent de stocker de manière



unitaire (une seule fois dans la table) des informations qui apparaissent souvent dans les autres tables (le rang hiérarchique "genre" est référencé 694 fois dans la table "ergotaxon", mais il n'est stocké qu'une seule fois dans la table "rang\_hierarchique"). La table "mode" stocke à l'heure actuelle un qualificatif qui renseigne sur l'usage d'un nomen, à savoir s'il est oublié, peu utilisé, utilisé seulement dans des publications systématiques ou également en dehors de la systématique. Une prochaine étape de travail sera d'ajouter à cette table un attribut (ou une colonne) "ref\_zoonompub" afin de stocker des références bibliographiques appuyant l'adjectif rattaché à tel nomen. La table "qualification\_nomen\_fonction\_autre\_nomen" permet de stocker les liens qui peuvent exister entre nomina (tel nomen est le nom original remplacé par tel nomen nouveau; tel nomen est l'orthographe subséquente de tel autre).

#### ALGORITHME

Lorsque l'utilisateur soumet une requête en indiquant un nomen de la série-classe, le prototype exécute dans l'ordre les étapes suivantes:

- (1) rechercher le nomen saisi par l'utilisateur dans la base de données;
- (2) vérifier dans la table "qualification\_nomen\_autre\_nomen" (fig. 5) si le nomen recherché est soit (a) un nouveau protonyme, soit (b) l'archéonyme (ou nomen ancien) d'un autre nomen (son néonyme), soit (c) le néonyme (ou *nomen novum*), introduit comme nomen de remplacement d'un nomen déjà existant (son archéonyme),
- (3) vérifier dans la table "qualification\_nomen\_autre\_nomen" si le nomen recherché est une orthographe subséquente à l'originale et indiquer l'originale, sinon c'est l'orthographe originale;
- (4) afficher le nomen de rang supérieur auquel le nomen recherché était subordonné dans la protaxinomie,
- (5) afficher pour le nomen de rang supérieur tous les nomina qui lui sont immédiatement subordonnés dans la protaxinomie;
- (6) afficher pour le nomen recherche tous les nomina qui lui sont subordonnés dans la protaxinomie;
- (7) afficher le ou les genres originellement inclus grâce à la table "listes\_genres\_originellement\_inclus";
- (8) à partir de l'ergotaxinomie actuelle, donner le nomen du taxon de rang le plus bas (soit le moins inclusif, ou *métrotaxon*) qui englobe le ou les genres inclus précédemment cités,
- (9) vérifier que les genres originellement exclus (information disponible dans la table "listes\_genres\_originellement\_exclus") sont à l'extérieur de l'extension taxinomique actuelle du métrotaxon: si des genres originellement exclus sont aujourd'hui inclus, ils seront indiqués sous l'appellation d'*intragenera*,
- (a) si l'extension taxinomique actuelle du métrotaxon correspondant au nomen recherche est définissable par ses genres inclus et ses genres exclus, le nomen sera qualifié de *choronyme*,
- (b) si l'extension taxinomique actuelle du métrotaxon correspondant au nomen recherché n'est définissable que par ses genres inclus, le nomen sera qualifié de *néonyme*.

(10) si le nomen est un choronyme, afficher les genres exclus et le nomen du taxon de rang le plus haut (soit le plus inclusif, ou *orotaxon*) qui englobe les genres inclus et ecarte les genres exclus de son extension taxinomique;

(11) vérifier le statut actuel du nomen recherché en fonction de son usage;

(12) proposer soit (a) une validation du nomen, que ce soit (a) parce que l'étendue taxinomique initiale et actuelle correspondent; ou (b) parce que d'un point de vue usage il est protégé, soit (b) s'il n'entre dans aucun de ces deux cas, une invalidation du nomen recherché, avec indication du nomen aujourd'hui à utiliser si l'on veut définir l'extension taxinomique que le nomen recherché représentait à l'origine (soit le nomen du métrotaxon ou de l'orotaxon quand c'est possible).

Lorsqu'un nomen se voit validé selon les Règles ici utilisées, il faut le protéger. Sa taxinomie actuelle devient sa taxinomie de référence. Cette procédure de validation est envisagée à l'heure actuelle uniquement à la condition qu'il y ait un contrôle par un spécialiste du groupe donné. Mais ce changement de taxinomie (passage de la taxinomie actuelle à la taxinomie de référence) peut se faire automatiquement. L'application des Règles peut donc être automatisée, et le résultat univoque pour tout utilisateur.

La recherche d'un nomen qui n'est pas de la série-classe donne un certain nombre d'informations mais ne donne pour l'instant pas le nom de la série-classe à utiliser à l'heure actuelle. Cela fait partie des évolutions qui doivent être apportées au logiciel. En effet, un *type de recherche* est à l'heure actuelle considéré mais il y en a au moins deux: (1) l'utilisateur a une idée du nomen de la série-classe à utiliser et veut s'assurer que c'est bien celui-ci qui doit être utilisé aujourd'hui, (2) l'utilisateur souhaite savoir quel est le nomen de la série-classe à utiliser aujourd'hui pour le genre sur lequel il travaille.

La structure de la base permet, à l'aide de requêtes simples, de retrouver aisément les informations stockées. Ce qui a soulevé le plus de difficultés est le parcours de la taxinomie nécessaire afin de retrouver l'attribution taxinomique actuelle des genres inclus et exclus. La méthode déployée présente une complexité algorithmique (ou CA dans la suite du texte, il s'agit d'un indice de complexité pour un algorithme) de  $2nk$ , où  $n$  désigne le nombre de genres et  $k$  le nombre de niveaux à parcourir (de rangs qu'il faut remonter dans l'arbre taxinomique) pour atteindre le taxon qui englobe les genres considérés. L'algorithme crée un tableau T qui stocke chaque taxon surordonné au fur à mesure qu'on remonte la taxinomie (qu'on change de rang) et son nombre d'occurrences.

Ce parcours de l'arbre taxinomique pourrait encore être amélioré. En effet, actuellement, l'algorithme utilise donne des temps de réponses acceptables mais la croissance de la base peut alourdir le parcours de l'arbre. Une optimisation serait alors envisageable en reprenant les schémas algorithmiques de Schieber et Vishkin (GUSFIELD, 1997 181-195, CORMEN et al., 2001 476-497) et en associant leur méthode de référencement des nœuds avec la proposition de numérotation (LITTLE, 1964, HULL, 1966, HELPEL, 1991). La proposition de Schieber et Vishkin repose sur des arbres dichotomiques voire équilibrés ce qui n'est pas le cas des taxinomies qui sont ici à traiter. Les auteurs proposent une méthode qui revient à étiqueter chaque nœud de leur arbre dichotomique équilibré par un 0 ou un 1 et permet donc de référencer chaque feuille par un nombre unique binaire. En partant de la racine, le référencement de chaque nœud d'un arbre taxinomique donne par un numéro permettrait d'associer

chaque taxon à un numéro et, au sein de cette taxinomie, de retrouver rapidement le taxon le plus ou le moins inclusif qui englobe les taxons considérés. On perd l'avantage du binaire mais c'est une obligation car les arbres à traiter dans notre exemple ne sont pas dichotomiques ni équilibrés. Cette perte n'est pas problématique car on ne perd pas l'optimisation au sein d'une taxinomie, puisque chaque taxon aura un numéro (par exemple 16.24.34 65.45 et 16.24.34.75.25) et retrouver le taxon qui inclut nos deux taxons exemples s'en trouvera tout aussi rapide et évident (ce sera ici 16 24.34). Parce que les taxinomies peuvent énormément varier, un nomen aurait un numéro *N* par taxinomie, et dans le cas d'un arbre qui ne présente pas dans chaque branche la même structure hiérarchique (soit le même emboîtement de rangs), on conservera la même structure hiérarchique pour chaque nomen au sein d'une taxinomie. Si pour un nomen un rang n'est pas spécifié, il sera représenté dans son numéro par un 0. Le 0 serait alors défini comme ne pouvant être le taxon supérieur commun qui englobe les autres taxons recherchés.

### UN EXEMPLE DE REQUÊTE

Le prototype accessible à l'heure actuelle [<http://lully.snv.jussieu.fr/zusudra>] se présente comme un moteur de recherche (fig. 6). Une fenêtre d'aide renseigne sur les noms interrogables.

Soit l'exemple d'une requête avec le nomen **AMPHIBIA**. La figure 7 représente la réponse à cette requête, affichée selon ces grandes catégories (1) informations sur le nomen: date, auteur(s), orthographe originale ou subséquente, (2) informations sur la place taxinomique de ce nomen dans la publication originale (nomen surordonné, autres nomina subordonnés au même nomen surordonné, et nomina subordonnés au nomen considéré); (3) comparaison, pour ce nomen, entre sa protaxinomie et son ergotaxinomie actuelle, (4) informations sur son usage, et proposition de validation; (5) commentaire.

Reprenons la figure 2 et l'exemple du nomen porté par le noeud T4. Le nomen qui lui est surordonné est le nomen porté par le noeud T2. Un autre nomen subordonné au même nomen surordonné est le nomen porté par le noeud T3. Les nomina subordonnés au nomen considéré sont les nomina portés par les noeuds G4, G5 et G6.

Développons l'exemple illustré par la recherche du nomen **AMPHIBIA**. La fiche résultat donne donc le nomen, ainsi que son auteur et sa date de publication (c'est-à-dire, Linné, 1758). Elle indique qu'il s'agit du nomen remplacé par **SEBAE** Scopoli, 1777 puis par **CRYEROSOA** Hermann, 1783 et qu'il s'agit de l'orthographe originale. De la table "protaxon", est extraite la taxinomie dans la publication originale. À partir de la liste de genres originellement inclus, le nomen à utiliser pour caractériser cette extension (dans l'ergotaxinomie) est indiqué soit dans le cas présent **VERTEBRATA** Cuvier, 1800. Une liste de genres explicitement originellement exclus est disponible mais certains de ces genres sont aujourd'hui inclus dans le taxon. Cette limite n'est donc pas informative pour fixer le statut du nomen. Comme l'indique la fiche, ce nomen ne peut être défini que par ses genres inclus. N'ayant pas été utilisé depuis 1899 au sens où il a été défini initialement, il n'y a pas lieu de le protéger. La validation propose donc de le remplacer par **VERTEBRATA** Cuvier, 1800, qui répond à la condition ici informative



Fig. 6. – Vue de la page d'accueil de l'interface web.

et utilisable (inclure aujourd'hui les genres qui étaient inclus dans **AMPHIBIA** Linnaeus, 1758). Dans le prototype actuel, il n'est proposé à l'utilisateur qu'une seule ergotaxinomie. Dans le futur, il est prévu d'en avoir plusieurs. En effet, tout comme dans le *Code*, il n'est pas question ici de définir une préférence pour un courant de pensée mais de permettre d'obtenir aisément des informations, de façon à prendre la décision la plus objective, la plus utile et la plus informative.

## DISCUSSION ET CONCLUSION

Ce travail vise à la publication d'un formalisme (stockage, exploitation et échange de données nomenclaturales et taxinomiques) appuyé sur un logiciel multi-plateforme. Il démontre que l'automatisation et l'informatisation de Règles nomenclaturales sont possibles. En effet, la nomenclature supérieure est plus compliquée à gérer que la nomenclature inférieure, et il sera donc relativement aisé d'appliquer ce formalisme à cette dernière. On pourrait alors

- **Nomen:** AMPHIBIA
- **Author(s):** Linnaeus
- **Date:** 1758
- **New or old nomen:** original nomen that has been replaced by SEBAE Scopoli, 1777 and CRYEROZOA Hermann, 1783
- **Spelling:** original spelling
- **Original taxonomy:**
  - Surordinate taxon to the taxon that bears the nomen ANIMALIA Linnaeus, 1758
  - Subordinate taxa to the same surordinate taxon AVES Linnaeus, 1758 INSECTA Linnaeus, 1758 MAMMALIA Linnaeus, 1758 PISCES Linnaeus, 1758, VERMES Linnaeus, 1758
  - Taxa subordinate to the taxon that bears the nomen NANTES Linnaeus, 1758 REPTILIA Linnaeus, 1758, SERPENTES Linnaeus, 1758
- **List of the generic nomina, originally included in the extension of the taxon bearing the nomen: (1)** Acipenser Linnaeus, 1758 [ ] (16) Testudo Linnaeus, 1758
- **Current taxonomic allocation of those generic nomina, originally included in the extension of the taxon bearing the nomen:** Subphylum VERTEBRATA Cuvier, 1800
- **List of the generic nomina, originally excluded in the extension of the taxon bearing the nomen: (1)** Alauda Linnaeus, 1758, [ ] (153) Zeus Linnaeus, 1758
- **Current taxonomic allocation of those generic nomina, originally excluded in the extension of the taxon bearing the nomen:** Subphylum VERTEBRATA Cuvier, 1800
- **Definition of the nomen:** the nomen can be defined only by its originally included taxa
- **Current status of the nomen regarding usage:** nomen clearly mentioned as nomenclaturally available (in some cases as an available senior homonym making a junior homonym invalid) but never used as valid by any author and in any publication after 31 December 1899
- **Proposed status of nomen:** nomenclaturally invalid but available nomen according to the rules here used for the class-series nomina: validation of a junior homonym or of a neonym: subjective synonym of Subphylum VERTEBRATA Cuvier, 1800
- **Comments:**  
The conulegenera [ ] the senior homonym

Fig. 7. – Fiche-résultat pour le nomen AMPHIBIA Linnaeus, 1758

s'attendre à un élargissement du champ d'action de *Cock*, tout en gardant à l'esprit que cela doit demander un minimum de travail, en termes de compréhension pour les utilisateurs (pour lesquels le logiciel ne doit pas, autant que faire se peut, être une boîte noire), et en termes de mise à jour.

Dans la suite de ce travail, s'imposeront: (1) une étape de communication avec des utilisateurs pour plus de valeur ajoutée: le site [paleodb.org] constitue un bon exemple d'un service qui propose un niveau d'information paramétrable par l'utilisateur; (2) une étape de communication avec des développeurs pour un échange de stratégies.

Enfin, un atout supplémentaire au déploiement du logiciel serait de mettre en place une interface de saisie conviviale, pour permettre à des contributeurs l'ajout de nouvelles publications dans la base

## RÉSUMÉ

Le *Code International de Nomenclature Zoologique* légifère sur les noms scientifiques ou *nomina* de rangs compris entre la sous-espèce et la super-famille, c'est-à-dire la nomenclature dite "inférieure". La nomenclature "supérieure" est actuellement "informelle", n'étant régie par aucune Règle, ce qui conduit à une situation d'instabilité nomenclaturale et de confusion. Une formalisation de la nomenclature supérieure récemment proposée repose sur le contenu et les limites des taxons pour lesquels les *nomina* ont été proposés, par comparaison avec les contenus et limites des taxons actuellement reconnus. De plus, afin de ne pas bouleverser deux siècles et demi de tradition, les *nomina* en usage important, spécialement en dehors du domaine de la systématique, doivent être conservés. Ce travail présente une formalisation pour la représentation des données nomenclaturales et l'automatisation de l'application de Règles de nomenclature proposées sur ces données. Un prototype de logiciel a été développé. Son architecture repose sur (1) une base de données qui comporte les informations nomenclaturales; (2) un programme situé côté serveur, d'interrogation de la base de données et d'application des Règles nomenclaturales; (3) une interface web d'exploitation du programme. Le prototype de ce logiciel a été développé dans un premier temps pour les *nomina* du groupe des Amphibiens.

## RÉFÉRENCES BIBLIOGRAPHIQUES

- ANONYMI [International Commission on Zoological Nomenclature], 1999. *International code of zoological nomenclature*. Fourth edition. London, International trust for zoological nomenclature [i-xxix] + 1-306
- ALONSO-ZARAZAGA, M. A., 2005. Nomenclature of higher taxa: a new approach. *Bulletin of Zoological Nomenclature*, **62** (4) 189-199
- BERENDSOHN, W. G., 1995. - The concept of "potential taxa" in databases. *Taxon*, **44**, 207-212
- , 1997. A taxonomic information model for botanical databases: the IOPI model. *Taxon*, **46** 283-309
- BERENDSOHN, W. G., ANAGNOSTOPOULOS, A., HAGEDORN, G., JAKUPOVIC, J., NIMIS, P. L., VALDES, B., GUNTCH, A., PANKHURST, R. J. & WHELI, R. J., 1999. A comprehensive reference model for biological collections and surveys. *Taxon*, **48** 511-562
- CHRISTIANSEN, T. & TORRINGTON, N., 1999. - *Perl in action*. Paris, O'Reilly [i-xxvii] + 1-972
- CORMIN, T. H., LEISERSON, C. E., RIVEST, R. L. & SHIN, C., 2001. - *Introduction to algorithms*. London, MIT Press: 1-1180

- DARLU, P & TASSY, P, 1993 *Reconstruction phylogénétique concepts et méthodes* Paris, Masson, Collection Biologie Théorique. 1-245
- DUBOIS, A, 2000 Synonymies and related lists in zoology general proposals, with examples in herpetology. *Dumerilia*, **4** (2): 33-98.
- 2004. – The higher nomenclature of recent amphibians. *Alytes*, **22** (1-2) 1-14.
- 2005a – Comment nommer les taxons supérieurs en zoologie et en botanique ? *Bios, stema*, **23** 3-10
- 2005b – Propositions pour l'incorporation des nomina de taxons de rang supérieur dans le *Code International de Nomenclature Zoologique*. *Biosystema*, **23**: 73-96
- 2005c – Proposed Rules for the incorporation of nomina of higher-ranked zoological taxa in the *International Code of Zoological Nomenclature* 1. Some general questions, concepts and terms of biological nomenclature. *Zoosystema*, **27** (2): 365-426
- 2005d – Proposals for the incorporation of nomina of higher-ranked taxa into the Code. *Bulletin of zoological Nomenclature*, **62** (4): 200-209
- 2006a – Proposed Rules for the incorporation of nomina of higher-ranked zoological taxa in the *International Code of Zoological Nomenclature* 2. The proposed Rules and their rationale. *Zoosystema*, **28** (1): 165-258.
- 2006b – New proposals for naming lower-ranked taxa within the frame of the *International Code of Zoological Nomenclature* *Comptes rendus Biologies*, **329** (10), 823-840.
- 2006c. – Incorporation of nomina of higher-ranked taxa into the *International Code of Zoological Nomenclature*: some basic questions. *Zootaxa*, **1337**: 1-37.
- DUBOIS, A & ÖHLER, A., 1997 Early scientific names of Amphibia Anura I Introduction *Bulletin du Muséum national d'Histoire naturelle*, (4) **18** (3-4): 297-320
- GRAY, J., O'NEIL, P & O'NEIL, E., 2000 - *Database principles, programming, performance* Second edition. San Francisco, Morgan Kaufmann Publishers: 1-870
- GUSFIELD, D., 1997 *Algorithms on strings, trees and sequences. computer science and computational biology*. Cambridge, Cambridge University Press: 1-534.
- HEPPELL, D. 1991 Names without number? In: D. L. HAKSWORTH (ed.), *Improving the stability of names: needs and options*, Königstein, Koeltz: 191-196.
- HULL, D. L., 1966. – Phylogenetic nomenclature. *Systematic Zoology*, **15** (1): 14-17.
- KNOX, E. 1998 The use of hierarchies as organizational models in systematics. *Biological Journal of the Linnean Society*, **63** 1-49.
- LAURENTI, J. N., 1768 *Specimen medicum, exhibens synopsin Reptilium emendatam cum experimentis circa venena et antidota Reptilium austriacorum* Viennae, Joan Thom Nob de Trattnern [3-n] + 1-215, pl. 1-5.
- LARIN, M & CANTINO, P. D., 2004 - First International Phylogenetic Nomenclature Meeting a report *Zoologica scripta*, **33**, 475-479
- LECOINRE, G & LE GUYADER, H., 2001 - *Classification phylogénétique du vivant* Paris, Belin 1-543 + [i-xxx]
- LINNAEUS, C., 1758 *Systema Naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus differentiis, synonymis, locis* Editio decima, reformata Tomus I Holmiae, Laurentii Salvii. [i-iv] + 1-824
- LITTLE, F. J. J. 1964 The need for a uniform system of biological nomenclature. *Systematic Zoology*, **13** (2): 191-194
- MAYR, E & ASHOCK, K. 1991 - *Principles of systematic zoology* Second edition New York, McGraw-Hill i-xx + 1-475
- SIMPSON, G. G., 1940 Types in modern taxonomy *American Journal of Science* **238** 413-431
- , 1961 *Principles of animal taxonomy* New York, Columbia Press University [i-xi] + 1-247
- TRUBB, L & CLOUTIER, C., 1991 A phylogenetic investigation of the inter- and intrarelationships of the Lissamphibia (Amphibia Temnospondyli) In: H.-P. SCHUBERTZ & L. TRUBB (ed.), *Origins of the higher groups of tetrapods: controversy and consensus*, Ithaca, New York: Cornell Univ Press 223-313
- TUDGE, C., 2000. – *The variety of life* Oxford, Oxford University Press [i-xv] + 1-684
- WALL, L., CHRISTIANSEN, T & ORWANT, J., 2001. – *Programmation en Perl*. Paris, O'Reilly [i-xxvii] + 1-1045

- WORSLEY, J. C. & DRAKE, J. D., 2002 - *Practical Postgre:SQL*. Paris, O'Reilly: [i-xiv] + 1-619
- ZHONG, Y., JUNG, S., PRAMANIK, S. & BLAMAN, J. H., 1996 Data model and comparison query methods for interacting classifications in taxonomic databases. *Taxon*, **45**: 223-241.
- ZHONG, Y., LUO, Y., SAKTI, P. & BLAMAN, J. H., 1999 HICLAS: a taxonomic database system for displaying and comparing biological classification and phylogenetic trees. *Bioinformatics*, **15** (2). 149-156

*Corresponding editor:* Annemarie OHLER.